DOCUMENT RESUME

ED 277 722 TM 860 733

AUTHOR Borg, Walter R.
TITLE How Did Methusilah Get into This? Tribulations of a Developer.
PUB DATE 3 Apr 85
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Associations (66th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Comparative Testing; *Content Validity; Educational Testing; Inservice Teacher Education; Intermediate Grades; Standardized Tests; *Test Construction; *Testing Problems; Test Validity; Time on Task; *Units of Study
IDENTIFIERS *Alternatives To Standardized Testing

ABSTRACT

Two approaches to testing student achievement are discussed: (1) traditional standardized tests; and (2) testing achievement on an instructional unit. Content validity is a major problem with standardized achievement tests. Even though test publishers typically consider a number of curriculum guides and textbooks in producing a test which is reasonably well suited to most schools' curricula, the curricular content in schools and classrooms is not consistent. For example, research has indicated wide variations in the amount of classroom time devoted to elementary school mathematics. The advantages of standardized tests are familiarity, simplicity, and economy. The proposed alternative involves the teaching of a unit of instruction, followed by an achievement test which measures only those concepts contained in the unit of study. A field test was conducted with intermediate school teachers and a week-long social studies unit on persons who had lived to be over 100. Although this approach required the development of the unit and the test, as well as the teachers' cooperation, it had high content validity. Observation indicated that 39 of 40 teachers followed the content of the instructional unit. (GDC)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED277722

# How Did Methuselah Get Into This? Tribulations of a Developer*

by

Walter R. Borg

UTAH STATE UNIVERSITY

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. Borg

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

## Abstract

Two approaches to the measurement of student achievement as a means of evaluating educational programs are discussed. The usual approach is to use standardized achievement tests. Several problems that arise from using these measures are discussed. The author proposes instead the development of special instructional units, not part of the regular curriculum, and measuring student achievement on the content of these units. This approach has several advantages. Most importantly it provides much better control of important variables and largely solves the content validity problem.

---

The goal of many of the educational programs that have been developed and evaluated in recent years has been to increase learner achievement. The usual method of evaluating such programs has been to administer a standardized achievement test to students who have been exposed to the program. However, determining the degree to which student achievement gains on such a test can be attributed to an instructional program is much more difficult than it appears. In this paper I would like to compare two approaches to the measurement of student achievement as a means of evaluating educational programs and look at some of the advantages and pitfalls involved in these approaches.

---

*Paper presented at the annual meeting of the American Educational Research Assn., Chicago, April 3, 1985.

WRB/E-1



TM 860133

Using Standardized Achievement Tests

First, lets take a look at the usual approach of evaluating student achievement through the use of standardized achievement measures. Suppose, for example, that we have developed a peer tutoring program and wish to evaluate this program by comparing the arithmetic gains of sixth grade students who have been exposed to peer tutoring as opposed to a comparable control group who have been exposed only to the conventional mathematics curriculum. The major problem that we face in using a standardized achievement measure to compare achievement outcomes of the experimental and control groups is finding a standardized test with satisfactory content validity. Even in an area as structured as arithmetic, different textbooks, and state and local curriculum guides differ considerably in the specific content taught, the time when it is taught, and the sequence that is followed. Test publishers typically attempt to establish content validity by studying a collection of curriculum guides, state courses of study, widely used textbooks and other related curriculum materials and then develop test items that measure a sample of the content most frequently found in these sources. This approach is aimed at producing a test that is reasonably well suited to the curricula of most local schools. It is clear, however, that the content objectives to be measured on a given standardized test will fit some schools much better than others. Although all major test publishers use a similar approach to establish content validity, the actual objectives measured can differ considerably from test to test. Even in mathematics where the curriculum tends to be less variable than most school subjects, researchers have found striking differences in the topics covered by popular standardized achievement tests (Freeman et. al., 1979; Floden et. al., 1978). Therefore, a program evaluation that cuts across several schools or school districts is

likely to employ a standardized achievement measure that fits the curriculum in some schools better than others, thus introducing an error when the achievement scores obtained are used as a dependent variable.

Even within the same school and at the same grade level it has been found that different teachers may differ considerably in the amount of time they devote to different school subjects and the specific content that they cover. The author recently conducted an exhaustive review of the literature on time allocation to different subject areas (Borg, 1980). All studies located showed very large differences in the time allocated to different subjects by different schools. The Beginning Teacher Evaluation Study provides the most detailed information that we have found on this topic. In this study, for example, it was found that the average amount of time allocated to mathematics in second grade classes varied from about twenty-five minutes a day to around sixty minutes per day. In studying time allocation for specific arithmatic topics, one second grade class was found in which the average student received nine minutes of instruction over the whole school year in the arithmetic associated with the use of money. This figure can be constrasted with classes where the average second grader was allocated three hundred and fifteen minutes per school year on curriculum content related to money. In another example, some fifth grade classes received less than a thousand minutes of instruction in reading comprehension for the school year, while other classes allocated almost 5,000 minutes to instruction related to reading comprehension during the school year (Fisher, C.W. et. al., 1978).

Another problem with using standardized achievement tests is that these are essentially survey instruments and as a result measure only a small sample of the curriculum content covered in the typical school. Furthermore, those concepts covered are typically measured by very few items.

When all of these deficiencies are considered it may be seen that the use of standardized achievement tests provides a very weak basis for evaluating the effectiveness of educational programs. The main advantage of using standardized achievement measures is that most people in education are familiar with these instruments and may be able to relate results (although often incorrectly) to their own educational setting. Another advantage of the standardized achievement measures is that they are relatively cheap and easy to use as compared to alternate methods of measuring achievment to evaluate educational products.

Testing Achievement on an Instructional Unit

Let us now consider an alternative approach to evaluating the achievment outcomes of educational programs that you have developed. This approach involves the development of a unit of instruction that all teachers who participate in the evaluation will teach in their classes. An achievement test is developed which measures only those concepts contained in this unit of instruction. Let me illustrate this method with a unit of instruction that we developed for use in one of our field tests. We had developed a program designed to train inservice teachers to emit twelve low inference verbal behaviors that were hypothesized to relate to pupil achievement. Examples of the behaviors covered in our program are "using voice modulation to emphasize main points", "paraphrasing important content from pupil responses," and "cueing important points in the discussion or recitation" with phrases such as, "This is important", or "Be sure to remember this." As part of our summative evaluation of this program we carried out a field test in which twenty-five intermediate grade elementary school teachers were trained in the program while a comparable control group of fifteen teachers did not receive the training. We had two criteria for evaluating the success of the program.

The first was concerned with the changes in teacher behavior, i.e., the degree to which the experimental group teachers employed the twelve verbal skills in their classrooms and, secondly, differences in achievement between students in the experimental versus the control classrooms.

The instructional unit developed was a social studies unit based on an article from National Geographic magazine. All participating teachers were asked to study the unit and teach it fifty minutes per day for four consecutive days. The achievement measure was then administered on Friday. The material given to the teacher included a copy of the article, a reading lesson based on the article, a short related article, a set of 35mm color slides of photographs from the article, and brief lesson plans the teachers could follow in teaching each lesson. The article, entitled, "Everyday is a Gift When You Are Over One Hundred," described the visit of a scientist to four areas in the world where a relatively large number of persons lived to the age of one hundred and beyond. Each of the areas was described along with the life style, diet, etc., of the aged inhabitants.

In selecting material for the development of a content unit several things should be considered. The content of the unit should be at a high interest level, and should fit the maturity level of the students. Furthermore, the content should be perceived by the teachers and other school officials as being relevant and worthwhile for the children to learn. Another requirement is that the content of the unit should not be something taught in the regular school curriculum and preferably should be a topic about which the child is unlikely to learn incidently from out-of-school sources.

Using a specially developed instructional unit as a basis for measuring student achievement has a number of disadvantages. First, the development of such a unit is in itself a fairly difficult task. Furthermore, after the unit

has been developed it is necessary to develop an achievment measure that covers all of the main concepts in the unit.

It is also necessary to obtain teacher cooperation to teach the unit. Unless all teachers teach the content of the unit, the achievement test you have developed will lack content validity. The most serious disadvantage of using such instructional units is that as a rule they must be taught over a fairly short time span. As the length of such units increases, the problems involved in getting the cooperation of the school administration to teach the unit and of getting teachers to carefully follow the unit content becomes increasingly more difficult.

Another problem emerges because of the amount of structure that must be imposed on teaching units of this type. In our study the goal of the instructional unit was to ensure that participating teachers taught essentially the same content and devoted the same amount of time to the teaching. Imposing these controls, of course, greatly reduces many of the errors inherent in evaluations that employ standardized achievement tests. However, the very structure that reduces errors also may tend to change or inhibit the teaching style of participating teachers. For example, relatively weak teachers who never prepare lesson plans of their own, may teach more effectively than usual if they are given well thought out lesson plans to follow as part of the instructional unit.

However, using instructional units also offers the evaluator some very important advantages. Perhaps the most important is that the achievement measure developed, since it covers virtually every major concept in the unit and nothing else, has extremely high content validity. Thus, this approach overcomes the main deficiency of using standardized achievement tests to measure the learning outcomes. Another advantage mentioned earlier is that by

using a carefully structured instructional unit the investigator has much better control over what is taught and how much time the teacher devotes to the subject. In our study, we were interested in keeping the content of the instruction constant so that variations in the use of the teaching skills which had been taught in our training program could be related to the achievement of students. It is doubtful that such relationships would often emerge if standardized achievement tests were used since the many errors inherent in that approach would almost surely mask the small achievement differences that are likely to be attributable to teacher use of a small number of specific instructional skills.

Another advantage of using the instructional unit is that since these units tend to be short, it is possible to use observers to determine the degree to which teachers are using the instructional skills and also to determine the teachers' adherence to the content of the instructional unit. In our evaluation an observer was present for two of the four 50-minute instructional periods. The teachers were aware that the observers would be coming but did not know which two periods they would be observed.

Before giving the teachers the packet of materials containing the instructional unit we pointed out to them the importance of following the unit closely since the achievement test given to their students at the end of the week would be based entirely on the content of the instructional unit. We pointed out that instructional time devoted to other topics, even topics that were relevant and worthwhile, would probably result in their students receiving lower scores on the achievement test since none of this peripheral material would be covered by the achievement test items.

The title of my paper, "How Did Methuselah Get Into This?" comes from our experience in using this instructional unit. On the second day of

instruction one of the observers came to see me and said that one of the teachers whom she had observed, although lecturing on the topic of old age, had not covered any of the material in the instructional unit. I went to observe this teacher on the following day and sure enough he spent the entire fifty minutes talking about Old Testament characters, including Mathusala, who had lived to very advanced ages. At the end of the lesson, the teacher who was obviously proud of himself for being so creative came to the back of the room and asked me how I liked his presentation. I told him I had found it very interesting and left. He was of course classified as an outlier and all of the data on his performance and his student's achievement had to be dropped from the study. The moral of this story is, of course, that the researcher should never underestimate the capacity of the human subject to screw-up the research process. Of the forty teachers involved in the field test this was the only one who deviated significantly from the content in the instructional unit. The use of the observers not only provided considerable evidence that the teachers had in fact followed the research procedure, but located this teacher, who if left in the study would have distorted the results since his students barely averaged the chance score on our achievement measure.

References

Borg, W. R. (1980). Time and School Learning in C. Denham and A. Lieberman (Eds.), Time to Learn. Washington, D.C.: National Institute of Education.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., Dishaw, M. M., & Moore, J. E. (1978). Teaching and Learning in the Elementary School: A Summary of the Beginning Teacher Evaluation Study. San Francisco: Far West Laboratory for Educational Research and Development.

Freeman, D., Kuhs, T., Knapper, L., & Porter, A. (1979). A closer look at standardized tests (Research Series No. 53). East Lansing, MI: Michigan State University, Institute for Research on Teaching (ERIC Document Reproduction Service No. ED 179581).

Floden, R. E., Porter, A. C., Schmidt, W. H., & Freeman, D. J. (1978). Don't they all measure the same thing? Consequences of selecting standardized tests. East Lansing, MI: Michigan State University, National Institute for Research on Teaching (ERIC Document Reproduction Service No. ED 167632).